Google

Company Profile

Google, Inc., is one of the most innovative companies on the Web. Its unique Internet search engine technology and infrastructure makes fast work of even the most laborious Web searches. Founded in 1998 by Stanford University PhD candidates Larry Page and Sergey Brin, Google has quickly and quietly emerged as one of the most popular search engines on the Web.

Despite having a limited advertising and marketing budget, (www.google.com and its co-branded partners) fields more million searches every day—up from just 500,000 per day! Currently the search engine of choice for Netscape’s popular Web site, the company also offers free and fee-based Goo Web searches to third-party Web sites.

Opportunity: Searches Made Fast

If Google’s business could be summed up in one word, it would be focus. While Web search behemoths like Yahoo!* and AltaVista and roll out e-mail, chat and other services, Google refines and extends its core search technology. The single-mindedness is starting to pay dividends in the form of lucrative OEM deals, escalating usage, and excellent word-of-mouth buzz.

A privately held company, Google earns money by licensing technology to customers like Netscape NetCenter*, Red Hat, Washington Post and Virgin Net* as well as by displaying its search results pages. True to the company’s mission of delivering clean searches, its ads consist of simple text.

One thing is certain, if you want your search results fast, you can’t do much worse than Google’s high-speed service. Performance measurements by independent publications consistently show Google to produce the fastest searches on the Web. What’s more, Google’s unique approach to assessing the value of results—by the number of links pointing to a site—seems to produce eerily relevant returns. The result: ever-increasing trust in its ability to deliver what users want.

In fact, Google was named to Time magazine’s "Best Cybersite 1999" list and ranked first in Time Digital magazine’s "Top Web Sites" list.
2000. Google also received a 1999 PC Magazine Technic Award. The search engine also twice ranked first overall as the leading search and portal sites in surveys of user satisfaction conducted by NPD Online Research in the last quarter and first quarter of 2000. Nevertheless, as Google has won customers, it has had to contend with breakneck growth.

Of course, all the accolades won't mean much if Google bases its business on expensive proprietary hardware. Jim Reese, chief operations engineer at Google, says that Intel® Architecture servers and a Linux* operating system fit Google's business plan. "With Linux* on the Intel® platform, there is no better price/performance we could find. When all is said and done, that is the bottom line."

**Solution: Off the Shelf, Vastly**

Google may be the ultimate scale-out proposition. The company employs redundant arrays of inexpensive servers (RAIS) to tap into an index between one and two terabytes (1,000 gigabytes) in size and consisting of over 200 million Web pages. To deliver whip-fast search results from such a massive database, Google employs a vast server farm of more than 3,500 uniprocessor servers based on Intel Architecture processors and linked using Intel PRO/100 Ethernet adapters. The servers are outfitted with 1GB of RAM and run the Linux operating system and a suite of built applications.

Google's application makes expensive proprietary hardware says Reese. "We are not like a transaction-based e-Comm where it makes sense to spend a whole lot of money on storage area network. We architected our solution to be scalable by using smaller servers that are multiply redundant and very fast through load balancing. Also it makes us very fault tolerant to lose a whole cluster or clusters, and we'll still be fine."

In Google's environment, disk I/O performance is an overriding factor, yet the cost of high-end SCSI disk subsystems is prohibitive. So the company standardized around inexpensive IDE technology, outfitting each server with a pair of internal disks storing either 22GB or 40GB apiece. "We did a lot of [performance comparisons] early on, and we found that for the best price/performance we would set up two IDE hard disks, each on a separate contr

Reese.

Google's massive search index is distributed and mirrored approximately 7,000 individual disk drives, enabling custom load-balancing software to point queries at the most-available disk subsystems. Network connectivity is enabled through PRO/100 Ethernet adapters at the servers linked to a gigabit backbone. The result: Peerless search response.
By employing redundant arrays of inexpensive servers (RA enables a vastly scalable, economical, and robust infrastructure that can adapt to surging demand. The company estimates that approximately 30 new servers a day to its farm, just to keep up with mounting demand. The servers are split among a pair of co-location facilities in the San Francisco Bay Area and a third facility on the East Coast. The company is also considering facilities in Asia as a way to cut down the latency for searches performed by users in Asia.

**Tailoring Solutions**

With 10,000 servers expected to be installed by the end of the year, co-location costs are an overriding factor. From the beginning, the company has worked with several OEMs to obtain uniquely compact servers and cabinets. The company says it can fit 80 servers in a space 7 feet tall by 2 feet wide and 2.5 feet deep. Hardware is sourced from numerous vendors to ensure timely delivery. Rackable Systems, an OEM specializing in compact servers, builds the systems.

Of course, managing such a large farm of servers presents its own challenge. The company crafted its own solutions for operational needs, such as remote management and load balancing. Google built its own software, including remote management consoles and load balancing engines. It has also developed a streamlined method for configuring servers that enables new systems to be brought online quickly.

"We have basically made our machines modular. We start with basic hardware and the base Linux OS, and then we 'Google you will,' with the base Google code that any machine might have. "At that point any machine can be made into a Web server, because all the machines have identical configurations."

Google continues to refine its services as well. The Google feature added to search results enables visitors to deepen their searches by Web pages related to a specific search return. The "Feeling Lucky" button on each search page takes the user to the page of the top-ranked result in a search. Google's high speed and rapid system configuration process enables the company to keep its infrastructure as new services like these are deployed.

**Summary**

With over 3,500 Intel® Architecture servers already humming along in Google's co-location facilities, the question is how many more will the company end up installing? At this point, the sky seems to be the limit. Google says it has been growing at a rate of 25% per year and as more and more Web sites employ Google's free and fast Web search service, that rate could accelerate.

One thing is certain, the company is ready, willing and able to deploy additional Intel Architecture servers to bear on the challenges it has crafted a vastly scalable infrastructure that makes the company's unique and sophisticated search engine technology. As traffic on www.google.com continues to grow by leaps and bounds, the company will simply add more servers to maintain Google's edge as the fastest search engines on the Web.